

## ТЕОРЕТИЧЕСКИЕ И ИСТОРИЧЕСКИЕ ПРОБЛЕМЫ ФРАЗЕОЛОГИИ

P. Grzybek (Graz)

### ZUM STATUS DER UNTERSUCHUNG VON SATZLÄNGEN IN DER SPRICHWORTFORSCHUNG – METHODOLOGISCHE VOR-BEMERKUNGEN

#### 0. Einleitung

Die Suche nach Prototypischem hat die Sprichwortforschung immer wieder auf verschiedene Art und Weise beschäftigt. Eines der Anliegen, welches dabei wiederholt vorgetragen wurde, ist die Frage nach der "typischen" Satzlänge von Sprichwörtern. Bei Versuchen, sich dieser Frage zu nähern, ist die Sprichwortforschung allerdings häufig nicht über den Status einer "symptomatischen" Beschreibung hinaus gekommen – eine systematische Beschreibungsebene hat sie in der Regel nicht erreicht.

Wir können an dieser Stelle nicht auf die Frage eingehen, unter welchen Bedingungen eine solche *systematische* Beschreibungsebene als erreicht angesehen werden kann, und wir müssen auch die mittlerweile fast 100 Jahre alte Geschichte der Satzlängenforschung ausblenden – diese Fragen werden an anderer Stelle ausführlicher zu behandeln sein (Grzybek 2000 a,b,c,d). Was im vorliegenden Text im Hinblick auf die genannte Problemstellung geleistet werden kann (und zunächst einmal geleistet werden muß), ist folgendes:

- eine Bestandsaufnahme, wo die im Rahmen der Parömiologie vorgenommenen Satzlängenforschungen stehen, d.h.: welche Arbeiten gibt es, was ist in ihnen untersucht worden, mit welchen Mitteln ist die Fragestellung bearbeitet worden (1.1.)

- ein Hinweis darauf, welche methodischen und methodologischen Probleme mit den bisherigen Untersuchungen und den in ihnen verwendeten Verfahren verbunden sind;

- die Durchführung kurzfristiger Reparaturmaßnahmen an den bisherigen Untersuchungen;

- die programmatische Formulierung weiterführender Untersuchungen.

#### 1.1. Estnisch (Krikmann 1967)

Eine der ersten Studien zur Satzlänge von Sprichwörtern dürfte Arvo Krikmann's (1967) in estnischer Sprache publizierter Aufsatz "Sprachstatistik estnischer Sprichwörter" sein. Diese Untersuchung wurde, vermutlich nicht zuletzt aufgrund sprachlicher Voraussetzungen, von der internationalen Sprichwortforschung – ungeachtet eines deutschen Resümées von Pentti Leino in *Proverbium 11* (1968) – praktisch nicht zur Kenntnis genommen, auch wenn sich diese Studie methodologisch in verschiedener Hinsicht als richtungsweisend hätte erweisen können. Krikmann hat drei verschiedene Stichproben estnischer Sprichwörter auf ihre in Worten gemessene Satzlänge untersucht:

1. 618 Sprichwörter aus den ältesten estnischen Quellen (17./18. Jhd.)

2. 3576 estnische Sprichwörter aus der Sammlung *Valimik eesto vanasõnu* von E. Normann (Tallinn, 1955);

3. 296 aktuell geläufige Sprichwörter, deren Bekanntheit zuvor in einer Erhebung mit ca. 20-30 Befragten erhoben worden war.

Die von ihm angeführten absoluten ( $f$ ) und relativen ( $f_r$ ) Häufigkeiten sind der Tab. 1 (vgl. Krikmann 1967, 130f.) zu entnehmen:

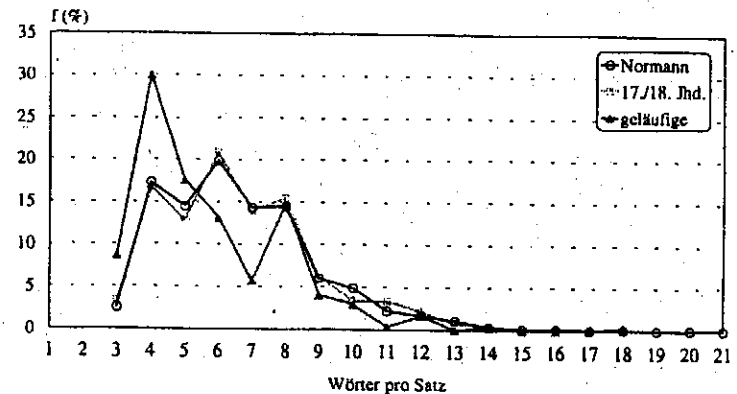
Tab. 1: Satzlängenhäufigkeit estnischer Sprichwörter nach Krikmann (1967)

Wörter pro Satz	17./18.Jhd.		Normann		geläufige	
	$f_i$	$f_{\%}$	$f_i$	$f_{\%}$	$f_i$	$f_{\%}$
1						
2						
3	21	3.40	93	2.60	26	8.78
4	103	16.67	619	17.31	89	30.07
5	80	12.94	518	14.49	52	17.57
6	129	20.87	710	19.85	39	13.18
7	86	13.92	511	14.29	17	5.74
8	96	15.53	518	14.49	44	14.86
9	39	6.31	218	6.10	12	4.05
10	21	3.40	175	4.89	9	3.04
11	21	3.40	81	2.27	1	0.34
12	14	2.27	62	1.73	5	1.69
13	4	0.65	40	1.12		0.00
14	3	0.49	13	0.36	1	0.34
15		0.00	5	0.14		0.00
16	1	0.16	5	0.14		0.00
17			2	0.06		0.00
18			4	0.11	1	0.34
19			0	0.00		
20			1	0.03		
21			1	0.03		
<b>Total</b>	<b>618</b>		<b>3576</b>		<b>296</b>	

Krikmann hat detaillierte Analysen insbesondere zu den 3576 Sprichwörtern der Normann-Sammlung durchgeführt. Krikmann hat eine Reihe zusätzlicher Berechnungen angestellt, auf die wir an dieser Stelle nicht detailliert eingehen können. So hat er u.a. auf einen Zusammenhang zwischen Sprichwortlänge (d.h. Satzlänge) und Wortlänge hingewiesen, demnach die Wörter, die ein Sprichwort bilden, um so kürzer sind, je länger das Sprichwort ist. Ungeachtet der Bedeutung dieser Beobachtungen werden wir diese

Beobachtung aus dem hier gegebenen Kontext ausblenden und an anderer Stelle systematisch aufgreifen müssen (Grzybek 2000d). Wir wollen uns zunächst auf die eigentliche Frage der Satzlänge beschränken, im Hinblick auf die Krikmann herausgestellt hat, daß ca. 4/5 des gesamten Materials (80.43%) im Intervall von 4-8 Wörtern pro Sprichwort liegen. Die durchschnittliche Länge der Sprichwörter hat Krikmann mit  $x = 6.57$  angegeben, wobei eine Nachberechnung seiner Daten ergibt, daß die mittlere Satzlänge de facto  $x = 6.62$  ausmacht.<sup>1</sup> Im Vergleich dazu beträgt die mittlere Länge der 296 geläufigen Sprichwörter  $x = 5.74$ , die der Sprichwörter aus dem 17./18. Jhd. hingegen  $x = 6.62$ . Ohne Mittelwertvergleiche oder weitere Analysen schlußfolgert Krikmann, daß die Satzlänge der geläufigen Sprichwörter sich deutlich von den übrigen unterscheidet, was sich auch in der graphischen Darstellung der Häufigkeitsverteilung zeige (Abb. 1, vgl. Krikmann 1967, 132).

Abb. 1: Häufigkeitsverteilung der Satzlängen estnischer Sprichwörter (Krikmann 1967)



### 1.2. Deutsch - Türkisch (Senaltan 1968)

In seiner Arbeit *Studien zur sprachlichen Gestalt der deutschen und türkischen Sprichwörter* hat Semahat Senaltan (1968, 6) 200 "ausgewählte deutsche Sprichwörter und ihre Sinnparallelen im Türkischen" im Hinblick auf die

Anzahl der in ihnen enthaltenen Wörter untersucht. Für das Deutsche traf der Autor eine nicht näher begründete Auswahl aus Wanders *Deutsches Sprichwörter-Lexikon*, für die türkischen Parallelsprichwörter diente die Sammlung *Türk atasözleri ve deyimleri* von F.F. Tülbentçi (Istanbul, 1963) als Quelle. Die Satzlängenhäufigkeiten hat Senaltan (1968, 43) als absolute Häufigkeiten ( $f_i$ ) dargestellt, die in Tab. 2 um die entsprechenden relativen Häufigkeiten ( $f_{\%}$ ) ergänzt wurden.

Tab. 2: Satzlängenhäufigkeit deutscher und türkischer Sprichwörter nach Senaltan (1968)

*Deutsch (n=200)*

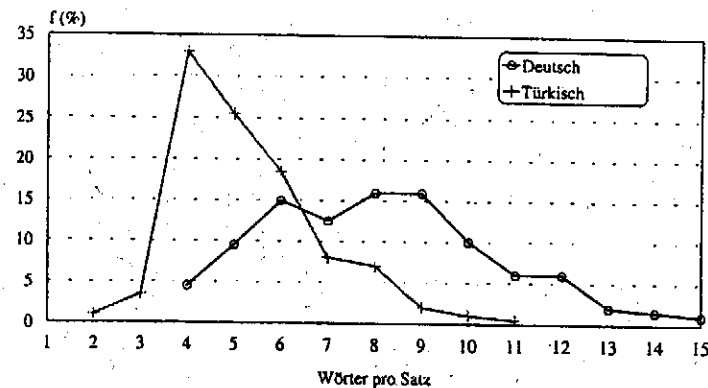
Wörter pro Satz	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
$f_i$				9	19	30	25	32	32	20	12	12	12	3	2
$f_{\%}$				4.50	9.50	15.00	12.50	16.00	16.00	10.00	6.00	6.00	2.00	1.50	1.00

*Türkisch (n=200)*

Wörter pro Satz	1	2	3	4	5	6	7	8	9	10	11
$f_i$		2	7	66	51	37	16	14	4	2	1
$f_{\%}$		1.00	3.50	33.00	25.50	18.50	8.00	7.00	2.00	1.00	0.50

Über die Erstellung einer primären Häufigkeitsliste hinaus sind von Senaltan keine statistischen Analysen durchgeführt worden. Allerdings wurde die Daten wie in Abb. 2 graphisch aufgearbeitet (vgl. Senaltan 1968, 43).

Abb. 2: Häufigkeitsverteilung der Satzlängen deutscher und türkischer Sprichwörter nach Senaltan (1968)



Nach Senaltan sind die auffälligen Satzlängenunterschiede zwischen den deutschen und türkischen Sprichwörtern auf sprachspezifische Faktoren zurückzuführen: Obwohl weder ein Mittelwertvergleich noch ein Vergleich mit anderen deutschen bzw. türkischen Texten (oder Sprichwortkorpora) geboten werden, geht der Autor von Längenunterschieden zwischen den deutschen und türkischen Sprichwörtern aus und führt die vermutete relative Kürze der türkischen Sätze auf den agglutinierenden Charakter des Türkischen zurück.

### 1.3. Deutsch – Russisch - Tschuwaschisch (Ivanov 1982)

Eine weitere Studie zur Untersuchung von 813 deutschen, 1000 tschuwaschischen, und 1087 russischen Sprichwörtern von Ivanov (1982).

Die 813 deutschen Sprichwörter stammten aus zwei Quellen: Aus dem *Немецко-русский фразеологический словарь* (Binovič/Grišin 1975) wurden alle enthaltenen 413 Sprichwörter übernommen, ebenso alle Einträge aus *400 немецких рифмованных пословиц и поговорок* von Petlevannyj/Malik (1980);

- die 1000 tschuwaschischen Sprichwörter stammten aus der Sammlung *Чувашские пословицы, поговорки и загадки* (Romanov 1960);

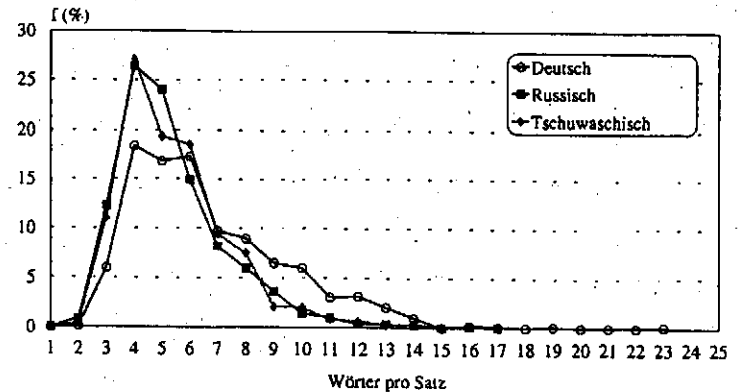
- die 1087 russischen Sprichwörter waren Äquivalente, die in den zuvor genannten Sammlungen angeführt werden: 813 waren in den beiden Sammlungen mit den deutschen Sprichwörtern aufgeführt, und 274 wurden in der tschuwaschischen Sammlung als russische Äquivalente angeführt (die übrigen 726 tschuwaschischen Sprichwörter wurden als wörtliche Übersetzungen bezeichnet). Die Daten werden von Ivanov (1982, 30) als Häufigkeitsliste mit den absoluten Werten wie in Tab. 3 dargestellt; zusätzlich führt Ivanov (1982, 32f.) die prozentualen Häufigkeiten sowie den jeweiligen Rang der betreffenden Satzlänge innerhalb der Stichprobe an. In Tab. 3 sind die entsprechenden Daten tabellarisch zusammengefaßt.

Tab. 3: Satzlängenhäufigkeit deutscher, russischer und tschuwaschischer Sprichwörter nach Ivanov (1982)

Wörter pro Satz	Deutsch		Russisch		Tschuwaschisch	
	$f_i$	$f_i$	$f_i$	$f_i$	$f_i$	$f_i$
1	0	0	0	0.00	0	0.00
2	1	0.12	10	0.92	4	0.40
3	49	6.03	134	12.33	110	11.00
4	150	18.45	287	26.40	271	27.10
5	137	16.85	261	24.01	194	19.40
6	141	17.34	162	14.90	186	18.60
7	79	9.72	89	8.19	94	9.40
8	73	8.98	65	5.98	75	7.50
9	53	6.52	40	3.68	21	2.10
10	49	6.03	16	1.47	21	2.10
11	25	3.08	11	1.01	8	0.80
12	26	3.20	4	0.37	7	0.70
13	17	2.09	4	0.37	3	0.30
14	8	0.98	3	0.28	3	0.30
15	0	0.00	0	0.00	1	0.10
16	2	0.25	1	0.09	1	0.10
17	1	0.12			1	0.10
19	1	0.12				
23	1	0.12				
Total	813		1087		1000	

Ivanov hat die von ihm erhaltenen Daten nach den Satzlengthen in Zweier-Klassen zusammengefaßt (2, 3-4, 5-6, 7-8, usw. Wörter pro Satz) und daraus eine Reihe vergleichender Aussagen abgeleitet. Die von ihm nicht dargebotene graphische Darstellungsform sei zum Zwecke der Vergleichbarkeit mit den anderen Ergebnissen wie in Abb. 3 nachgetragen.

Abb. 3: Häufigkeitsverteilung der Satzlengthen deutscher, russischer und tschuwaschischer Sprichwörter nach Ivanov (1982)



Aufgrund seiner Analysen gelangt der Autor zu verschiedenen Beobachtungen und Schlußfolgerungen, wobei die beiden folgenden im Hinblick auf die Frage der Satzlengthe die wichtigsten (weil allgemeinsten) sind:

1. Die durchschnittliche Satzlengthe ist für alle drei Stichproben unterschiedlich: Während sie für die russischen und tschuwaschischen Sprichwörter Ivanov (1982, 32) zufolge "in etwa gleich" sei ( $x = 5.32$  vs.  $x = 5.47$ ), stelle sie sich für die deutschen Sprichwörter als deutlich länger dar ( $x = 6.63$ ).

2. Für alle drei Stichproben seien Sprichwörter mit vier Wörtern die häufigsten, gefolgt von Sprichwörtern mit fünf, sechs oder sieben Wörtern (in unterschiedlicher Reihenfolge für die drei Sprachen).

Der Autor führt zwar für die jeweiligen Durchschnittslängen den Mittelwert an, nicht aber die entsprechenden Standardabweichungen, deren Kenntnis notwendig ist, um die beobachteten Mittelwertunterschiede auf ihre Signifikanz hin zu untersuchen (s.u.).

#### 1.4. Ungarisch - Russisch (Tóthné Litovkina 1990)

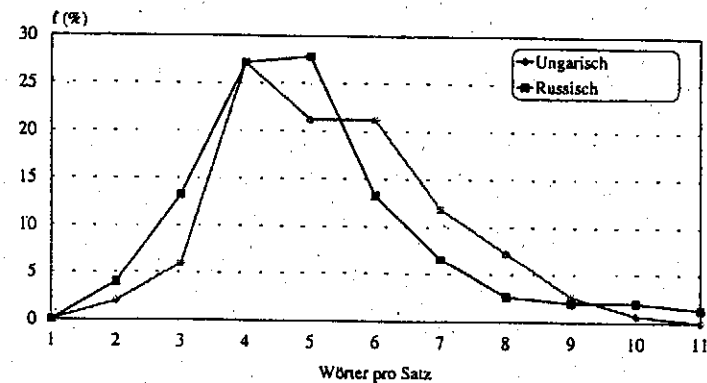
Eine weitere Untersuchung zur Satzlänge von Sprichwörtern stammt von Tóthné Litovkina (1990). Die Autorin hat die Satzlänge von 151 ihren Angaben zufolge bekannten ungarischen Sprichwörtern und deren 151 "vollständigen semantischen Äquivalenten" im Russischen untersucht. Der Autorin zufolge fallen 92% des ungarischen und 89% des russischen untersuchten Sprichwortmaterials in die Kategorie von 3-7 Wörtern pro Satz. Wie ein erster Blick auf das in Tab. 4 reproduzierte Datenmaterial zeigt, hat die Autorin sich bei der Angabe der Prozentzahlen geirrt: Tatsächlich fallen 88.08% des russischen ( $N=133$ ) und 87.42% des ungarischen Materials ( $N=132$ ) in den genannten Bereich; die von der Autorin erwähnten 92.05% der russischen Sprichwörter ( $N=139$ ) und 87.42% der ungarischen Sprichwörter ( $N=135$ ) hingegen betreffen den Bereich der Sätze mit 2-7 Wörtern.

Tab. 4: Satzlängenhäufigkeit ungarischer und russischer Sprichwörter nach Tóthné Litovkina (1990)

Wörter pro Satz	Ungarisch		Russisch	
	$f_i$	$f_{\%}$	$f_{abs}$	$f_{\%}$
1	0	0.00	0	0.00
2	3	1.99	6	3.97
3	9	5.96	20	13.25
4	41	27.15	41	27.15
5	32	21.19	42	27.81
6	32	21.19	20	13.25
7	18	11.92	10	6.62
8	11	7.28	4	2.65
9	4	2.65	3	1.99
10	1	0.66	3	1.99
11	0	0.00	2	1.32
Total	151		151	

Ungeachtet der abweichenden Prozentzahl scheint die Schlußfolgerung der Autorin auch durch die von ihr dargebotene (Tóthné Litovkina 1990, 251) graphische Darstellung (Abb. 4) unterstützt zu werden: Demnach stellt das Spektrum der Sprichwörter mit 2-7 bzw. 3-7 Wörtern pro Sprichwortsatz sowohl bei den russischen als auch bei den ungarischen Sprichwörtern offenbar eine prototypische Länge dar, was die Autorin als Indiz einer die Merkfähigkeit fördernden Eigenschaft interpretiert. Es werden im Hinblick auf die Beobachtungen bzw. Schlußfolgerungen der Autorin allerdings keine statistischen Analysen durchgeführt, die als Beleg ihrer Vermutungen dienen könnten.

Abb. 4: Häufigkeitsverteilung der Satzlängen ungarischer und russischer Sprichwörter nach Tóthné Litovkina (1990)



#### 1.5. Deutsch (Grzybek 1995)

Grzybek (1995) hat versucht, die Bekanntheit von Sprichwörtern als einen auf die Satzlänge einflussnehmenden Faktor herauszustellen. Einerseits auf der Grundlage der Bekanntheitsangaben der Sprichwörter aus Simrocks Sammlung *Deutsche Sprichwörter*, wie sie von Chlosta / Grzybek / Roos (1994) durch Befragung von 20 Muttersprachler(inne)n erarbeitet wurden, andererseits auf der Basis der in Wörtern gemessenen Satzlängen

dieser Sprichwörter, führte Grzybek (1995) gesonderte Satzlängenanalysen durch für:<sup>2</sup>

- a) das gesamte Korpus der 12.977 Sprichwörter;
- b) die 346 Sprichwörter, die einen Bekanntheitsgrad von > 50% aufwiesen (also mindestens 11 der 20 befragten Personen bekannt waren);
- c) die 723 Sprichwörter, die einen Bekanntheitsgrad von > 75% aufwiesen (also mindestens 16 der 20 befragten Personen bekannt waren);
- d) die 61 Sprichwörter, die 100% (also allen 20 befragten Personen) bekannt waren.

Übergeordnete Fragestellung der Untersuchung war, ob sich die durchschnittliche Satzlänge der Sprichwörter in Abhängigkeit vom Grad ihrer Bekanntheit verändert. Durchgeführt wurden einerseits Mittelwertberechnungen für die jeweiligen Satzlängen des Gesamtkorpus und der Sprichwörter in den drei unterschiedenen Bekanntheitsklassen, andererseits Häufigkeitsberechnungen, d.h. Analysen, wie häufig welche Satzlängen in den jeweiligen Klassen vorkommen. Im Hinblick auf die Mittelwertberechnungen wurden für die angeführten Bekanntheitsklassen jeweils die durchschnittlichen Satzlängen mit den dazugehörigen Standardabweichungen dargeboten. Die (korrigierten) Werte sind in Tab. 5 dargestellt.

Tab. 5: Absolute und prozentuale Häufigkeit der unterschiedenen Bekanntheitsklassen mit jeweiligem Mittelwert und Standardabweichung der Satzlängen nach Grzybek (1995)

Bekanntheit	N	in %	x	s
gesamt	12977	100	7.86	3.41
> 50%	723	5.57	6.60	2.42
> 75%	346	2.66	6.35	2.35
100%	61	0.47	6.16	2.37

Es wurden die Mittelwerte der Satzlängen in den unterschiedenen Bekanntheitsklassen berechnet und mit Hilfe von *t*-Tests miteinander (s.u.) verglichen. Dabei zeigte

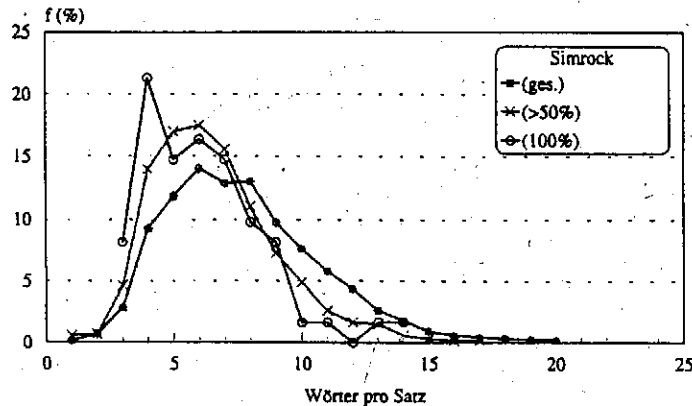
sich, daß sich die Durchschnittslänge der Sprichwörter des gesamten Korpus signifikant von der Durchschnittslänge aller anderen unterschiedenen Bekanntheitsklassen unterscheiden (so erreichte z.B. der Unterschied zwischen der Gesamtmenge zu den mehr als 50% bekannten Sprichwörtern ein hohes Signifikanzniveau:  $t_{FG=890} = 13.28$ ;  $p < 0.001$ ).

Berechnet wurden auch Mittelwertvergleiche zwischen den Ergebnissen der Satzlängenanalysen deutscher Sprichwörter von Senaltan, Ivanov und den drei unterschiedenen Bekanntheitsklassen. Da solche Vergleiche unten systematisch durchgeführt werden, kann an dieser Stelle auf eine Darstellung der Ergebnisse verzichtet werden.

Abgesehen von den Berechnungen der durchschnittlichen Satzlängen in den drei unterschiedenen Bekanntheitsklassen wurde versucht, über die Verteilung der Satzlängenhäufigkeiten in diesen verschiedenen zusammengefaßten Klassen Aufschluß über den Einfluß des Faktors Bekanntheit zu gewinnen. Es wurden zu diesem Zweck die absoluten Vorkommenshäufigkeiten der jeweiligen Satzlängen in prozentuale Häufigkeiten umgerechnet und die jeweiligen Satzlängenhäufigkeiten wie in Abb. 5 (vgl. Grzybek 1995, 213) graphisch dargestellt. Es zeigt sich, daß sich mit zunehmender Bekanntheit der Sprichwörter der Verlauf der Kurve ändert: sie wird steiler und verlagert sich nach links (der Exzess wird deutlicher positiv, die Kurve leptokurtisch); dies wurde als zusätzliches Indiz dafür angesehen, daß die durchschnittliche Satzlänge mit zunehmender Bekanntheit der Sprichwörter abnimmt. Um diese aufgrund des optischen Eindrucks gewonnene Annahme abzusichern, wurden jedoch keine entsprechenden statistischen Verfahren angewendet. Allerdings wurde darauf hingewiesen, daß es bei der Untersuchung des Einflusses von Bekanntheit auf Satzlänge der Sprichwörter sinnvoller wäre, keine Mittelwertvergleiche zwischen mehr oder weniger willkürlich festgelegten Bekanntheitsklassen durchzuführen, sondern im Falle eines linearen Zusammenhangs den Regressionskoeffizienten

(s.u.) zu bestimmen sowie einen Test auf Linearität durchzuführen, um eine etwaige Abweichung von der Linearität feststellen zu können.

Abb. 5: Häufigkeitsverteilung der Satzlängen deutscher Sprichwörter in drei unterschiedenen Bekanntheitsklassen nach Grzybek (1995)



## 2. Zusammenfassung und Programm

Die bisherigen Untersuchungen zur Satzlänge von Sprichwörtern unterscheiden sich zum Teil nicht unerheblich in den von ihnen verwendeten Methoden bzw. im Umgang mit dem verwendeten statistischen Inventar. Verwendung gefunden haben Mittelwertberechnungen, Darstellungen primärer Häufigkeitsverteilungen in Form von absoluten und/oder relativen Häufigkeitswerten sowie deren graphische Repräsentation. Positiv zu bewerten ist, daß in allen Untersuchungen die Daten in Form von primären Häufigkeitslisten mit den absoluten Vorkommnissen aufgeführt worden sind, so daß im Hinblick auf eine Reihe von Fragestellungen Re-Analysen, z.T. auch ergänzende bzw. weiterführende Analysen durchgeführt werden können. Insofern stellen die bisherigen Arbeiten wertvolles Datenmaterial zur Verfügung.

Andererseits gilt es nachdrücklich zu bemängeln, daß in den genannten Arbeiten zum Teil weitreichende Schlußfolgerungen gezogen werden, ohne daß dabei immer die notwendigen Erfordernisse im Umgang mit angemessenen statistischen Verfahren erfüllt worden wären: Dies betrifft beispielsweise die Tatsache, daß intuitive Mittelwertvergleiche ohne entsprechende Tests durchgeführt werden, oder daß aufgrund graphischer Repräsentationen optische Eindrücke als ausreichend angesehen werden, Schlußfolgerungen im Hinblick auf die Häufigkeitsverteilungen anzustellen usw. Die Validität verschiedener Folgerungen und Interpretationen muß damit nachhaltig in Frage gestellt werden.

Die Notwendigkeit, die Frage der Satzlänge in Sprichwörtern nicht nur systematisch anzugehen, sondern auch mit angemessenen statistischen Verfahren zu behandeln, liegt auf der Hand. Zur Behebung dieses Zustandes bietet sich folgendes Vorgehen an:

1. Als erster Schritt, die offenkundigen Desiderata der Satzlängenforschung innerhalb der Parömiologie zu behandeln, scheint es sinnvoll, die in den bisherigen Studien angewandten statistischen Verfahren kritisch zu überprüfen und, wo möglich, durch entsprechende Re-Analysen allfällige Unzulänglichkeiten in diesen Arbeiten zu bereinigen.

2. Ein zweiter Schritt hat dann darin zu bestehen, weitere, bislang nicht angewendete statistische und quantifizierende Verfahren perspektivisch vorzustellen, die eine Lösung der anstehenden Fragen herbeiführen können.

Während der erste, im folgenden als "Reparaturmaßnahmen" bezeichnete Schritt im Rahmen der vorliegenden Darstellung getan werden kann, läßt sich die im zweiten Schritt avisierte Lösung seit langem offener Fragen natürlich nicht in einem einzelnen Aufsatz erreichen; sie muß ihrerseits in mehreren Schritten herbeigeführt werden und kann deshalb hier nur abschließend programmatisch formuliert werden.

### 3. Reparaturmaßnahmen

#### 3.1. Datenpräsentation

Grundsätzlich sollten Daten zumindest in Form einer primären Häufigkeitsliste die absoluten Vorkommenshäufigkeiten ( $f_i$ ) wiedergeben. Geboten werden sollte ergänzend die relative ( $p_i$ ) bzw. prozentuale ( $p_{\%}$ ) Vorkommenshäufigkeit, d.h.

$$p_i = \frac{f_i}{N} \text{ bzw. } p_{\%} = \frac{f_i \cdot 100}{N}$$

Fakultativ sind weitere Angaben wie Rang, Summenhäufigkeit, relative Summenhäufigkeit u.a.m. – Angaben, die sich ansonsten bei Bedarf leicht nachträglich berechnen lassen. Um die Häufigkeitsverteilung der jeweiligen Stichprobe in einer Graphik anschaulich darzustellen, können entweder die absoluten oder die relativen Häufigkeiten herangezogen werden; beim Vergleich zweier Stichproben bieten sich bei der graphischen Darstellung die relativen Häufigkeiten an, um durch unterschiedliche Stichprobengrößen bewirkte optische Verzerrungen zu vermeiden.

Die hier vorgesehenen Vorschläge zur Darstellung von Daten sind oben (1.1.-1.5.) bereits vollzogen, so daß es hier keinerlei weiterer "Reparaturmaßnahmen" bedarf.

#### 3.2. Mittelwerte und Standardabweichungen bzw. Varianzen

Mittelwerte werden in der Regel angeführt, um die Mittelwerte verschiedener Stichproben miteinander vergleichbar zu machen. Mittelwertvergleiche sind aber nur sinnvoll, wenn Unterschiede in den entsprechenden Mittelwerten auch auf Signifikanz untersucht werden können. Dazu ist es notwendig, die jeweilige Standardabweichung bzw. Varianz zu kennen und bei entsprechenden Untersuchungen anzugeben. Rufen wir uns die entsprechenden Formeln ihrer Berechnung ins Gedächtnis:

Das *arithmetische Mittel* ( $\bar{x}$ ) berechnet sich als

$$(1) \quad \bar{x} = \frac{x_1 + x_2 + x_3 + \dots + x_N}{N} = \frac{1}{N} \sum_{i=1}^N x_i$$

bzw. bei gruppiertem Datenmaterial als

$$(1a) \quad \bar{x} = \frac{x_1 f_1 + x_2 f_2 + \dots + x_k f_k}{N} = \frac{1}{N} \sum_{i=1}^k x_i f_i$$

wobei  $N = \sum_{i=1}^k f_i$

Die *Varianz* ( $s^2$ ) berechnet sich nach

$$(2) \quad s^2 = \frac{1}{N-1} \sum_{i=1}^N (x_i - \bar{x})^2$$

bzw. bei gruppiertem Datenmaterial nach

$$(2a) \quad s^2 = \frac{1}{N-1} \sum_{i=1}^k (x_i - \bar{x})^2 f_i$$

Die *Standardabweichung* ( $s$ ) entspricht der Wurzel aus  $s^2$  bzw. der Wurzel aus dem Mittelwert der quadrierten Abweichungen:

$$(3) \quad s = \sqrt{\frac{\sum (x - \bar{x})^2}{N-1}}$$

Berechnet wird die Standardabweichung in der Regel nach (3a) oder (3b):

$$(3a) \quad s = \sqrt{\frac{\sum x_i^2 - \frac{(\sum x_i)^2}{N}}{N-1}} \quad (3b) \quad s = \sqrt{\frac{N \sum x_i^2 - (\sum x_i)^2}{N(N-1)}}$$



Auf der Grundlage von Mittelwert und Varianz bzw. Standardabweichung lassen sich Mittelwertvergleiche in Form des  $t$ -Tests durchführen. Der  $t$ -Test existiert in zwei Formen, dem Einstichproben- und dem Zweistichproben- $t$ -Test: Mit dem Einstichproben- $t$ -Test wird die Frage geprüft, ob der Mittelwert einer konkreten Stichprobe zufällig oder statistisch signifikant von einem vorgegebenen Mittelwert verschieden ist. Mit dem Zweistichproben- $t$ -Test wird geprüft, ob sich die Mittelwerte zweier verschiedener Stichproben unterscheiden. Geprüft wird die Nullhypothese, daß sich zwei Mittelwerte auf einem festgelegten Niveau nicht unterscheiden. Sind mehrere Mittelwerte- etwa aus Stichproben-Untergruppen – miteinander zu vergleichen, dann darf der  $t$ -Test nur nach Bonferoni-Korrektur oder einem äquivalenten Verfahren angewendet werden; alternativ wäre eine Varianzanalyse mit anschließenden post-hoc-Tests nach Scheffé zu durchzuführen.

Für die oben angegebenen Zwecke, d.h. für die mehrfach gestellte Frage, ob die Satzlänge verschiedener Stichproben von Sprichwörtern sich signifikant unterscheidet, ist also die Prüfung der Nullhypothese auf Gleichheit zweier Mittelwerte bei ungleichen (bzw. möglicherweise ungleichen) Varianzen durchzuführen, entsprechend wird Formel (4) verwendet:

$$(4) \quad t = \frac{|\bar{x}_1 - \bar{x}_2|}{\sqrt{\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}}}$$

Zur Überprüfung der Signifikanz von Mittelwertunterschieden ist der Vergleich in einer  $t$ -Tabelle (vgl. z.B. Sachs 1992, 210) notwendig. Hierbei müssen die Freiheitsgrade ( $v$ ,  $FG$ ) berücksichtigt werden, die nach (5) berechnet werden:

$$(5) \quad v = \frac{\left(\frac{s_1^2}{N_1} + \frac{s_2^2}{N_2}\right)^2}{\frac{\left(\frac{s_1^2}{N_1}\right)^2}{N_1 - 1} + \frac{\left(\frac{s_2^2}{N_2}\right)^2}{N_2 - 1}}$$

### 3.3. Unmittelbare Reparaturen

Wenn wir uns nach diesen allgemeinen Formulierungen nun noch einmal die oben referierten Untersuchungen zur Satzlänge von Sprichwörtern ins Gedächtnis rufen, so läßt sich folgendes sagen: In zwei der fünf Arbeiten (Senaltan, Tóthné Litovkina) sind überhaupt keine Mittelwerte angegeben; in den drei übrigen Arbeiten (Krikmann, Ivanov und Grzybek) werden zwar Mittelwerte angeführt, aber weder bei Krikmann noch bei Ivanov werden die dazugehörigen Standardabweichungen genannt – wie zu sehen war, ist dies aber unabdingbare Voraussetzung, um Größenunterschiede bei Mittelwerten deskriptiv einschätzen und allfällige Mittelwertunterschiede auf Signifikanz hin überprüfen zu können.

Da die primären Häufigkeiten in allen Untersuchungen aufgeführt sind, lassen sich die entsprechenden Berechnungen in Form von Re-Analysen anstellen; die Mittelwerte und Standardabweichungen aller dargestellten Studien zur Satzlänge in Sprichwörtern sind in Tab. 6 zusammenfassend dargestellt.

Tab. 6: Stichprobenumfang ( $N$ ), Mittelwerte ( $\bar{x}$ ) und Standardabweichungen ( $s$ ) der Untersuchungen von Satzlängen in Sprichwörtern

	$N$	$\bar{x}$	$s$
<b>Krikmann</b>			
estnisch (17./18. Jhd.)	628	6.62	2.25
estnisch (Normann)	3576	6.62	2.30
estnisch (gelaufte)	296	5.74	2.24
<b>Senaltan</b>			
deutsch	200	8.15	2.43
türkisch	200	5.29	1.54
<b>Ivanov</b>			
deutsch	813	6.63	2.75
russisch	1087	5.33	1.96
tschuwasch.	1000	5.47	2.02
<b>Tóthné Litovkina</b>			
russisch	151	4.95	1.76
ungarisch	151	5.36	1.58
<b>Grzybek</b>			
deutsch (Simrock)	12977	7.86	3.41
deutsch (Simrock >50%)	723	6.60	2.42
deutsch (Simrock 100%)	61	6.16	2.37

Aufgrund der in Tab. 6 angeführten Werte lassen sich nun eine Reihe von Nachberechnungen anstellen, die einerseits die von den Autor(inn)en selbst gezogenen Vergleiche und Schlußfolgerungen betreffen, andererseits auch Vergleiche zwischen den verschiedenen Studien erlauben. Durchführen lassen sich im Prinzip Mittelwertvergleiche zwischen allen erhaltenen Ergebnissen – vergleichen kann man eben letztlich praktisch alles mit allem. Sinnvoll erscheint jedoch eine Beschränkung auf zweierlei Typen von Vergleich:

(a) Vergleiche zwischen Stichproben, die von einem Autor erhoben wurden, und die sich auf Sprichwörter unterschiedlicher Sprachen beziehen: Deutsch – Türkisch (Senaltan), Deutsch – Russisch, Deutsch – Tschuwaschisch, Russisch – Tschuwaschisch (Ivanov), Russisch – Ungarisch (Tóthné Litovkina).

(b) Vergleiche zwischen Stichproben, die sich auf Sprichwörter ein und derselben Sprache beziehen; dies betrifft die Untersuchungen zu estnischen Sprichwörtern (Krikmann), zu russischen Sprichwörtern von Ivanov und Tóthné Litovkina sowie zu deutschen Sprichwörtern von Senaltan, Ivanov und Grzybek, wobei es sich bei Grzybek anbietet, beim Vergleich die einzelnen unterschiedenen Bekanntheitsklassen jeweils getrennt zu berechnen.

Durchgeführt wurden im Rahmen einer Re-Analyse entsprechende  $t$ -Tests. Dabei bleibt das Problem der (zumal unterschiedlichen) Stichprobengrößen hier nicht weiter verfolgt werden; in einem nächsten Schritt wäre jedenfalls unbedingt zusätzlich ein Maß der praktischen Signifikanz (wie z.B. der Eta-Koeffizient) zu berechnen.<sup>3</sup>

Tab. 7 gibt zunächst die Ergebnisse für die Vergleiche innerhalb der jeweiligen Studien wieder; auf dem 5%-Niveau signifikante Unterschiede ( $p < 0.05$ ) sind hier und im folgenden mit einem '\*' gekennzeichnet, auf dem 1%-Niveau ( $p < 0.01$ ) hochsignifikante Unterschiede mit '\*\*', auf dem 0.1%-Niveau hochsignifikante Unterschiede ( $p < 0.001$ ) mit '\*\*\*'.

Tab. 7: Mittelwertvergleiche zwischen den Stichproben mit Sprichwörtern verschiedener Sprachen

	Stichprobe 1	Stichprobe 2	$t$	$FG$	$p$
<b>Senaltan</b>	Deutsch	Türkisch	14.06	337	< 0.001***
<b>Ivanov</b>	Deutsch	Russisch	11.49	1394	< 0.001***
	Deutsch	Tschuwasch.	10.04	1452	< 0.001***
	Russisch	Tschuwasch.	1.61	2058	> 0.05
<b>Tóthné Litovkina</b>	Russisch	Ungarisch	2.13	297	< 0.05*

1. Für die Untersuchung von Senaltan (1968) stellt sich damit heraus, daß die von ihm beobachteten Satzlängenunterschiede sich in der Tat in einem hoch signifikant unterschiedlichen Mittelwert ( $p < 0.001$ ) ausdrücken.

2. Für die Untersuchung von Ivanov (1982) bestätigen sich ebenfalls die vom Autor angestellten Vermutungen, insofern sich nämlich die Unterschiede zwischen Deutsch und Russisch sowie zwischen Deutsch und Tschuwaschisch jeweils als hoch signifikant erweisen ( $p < 0.001$ ), der Unterschied zwischen Russisch und Tschuwaschisch hingegen kein Signifikanzniveau erreicht.

3. Für die Untersuchung von Tóthné Litovkina – die ja keine Mittelwertberechnungen vorgenommen hatte – stellt sich allerdings heraus, daß sich die Mittelwerte zwischen den von ihr untersuchten russischen und den ungarischen Sprichwörtern signifikant unterscheiden ( $p < 0.05$ ). Dieses Ergebnis wird durch die von der Autorin gemachte Beobachtung unterstützt, daß von den insgesamt 151 Sprichwortäquivalenten in 84 Fällen die ungarischen länger seien als die russischen, in 32 Fällen die russischen hingegen länger als die ungarischen, während in 35 Fällen die Länge identisch sei. – Wenn man sich vor dem Hintergrund des nunmehr festgestellten Mittelwertunterschieds daran erinnert, daß die Autorin (aufgrund der Häufigkeitsverteilung) eher eine Ähnlichkeit denn eine Divergenz zwischen den ungarischen und russischen Sprichwörtern suggerierte, so haben diese Beobachtungen methodologisch weitreichende Konsequenzen: sie führen einerseits zu der Frage der Aussagekraft von Mittelwerten überhaupt, und sie weisen andererseits auf die Notwendigkeit hin, der *Analyse von Häufigkeitsverteilungen* systematisch (und nicht nur aufgrund optischer Eindrücke, wie bei Tóthné Litovkina) nachzugehen – was an dieser Stelle freilich nicht in der notwendigen Ausführlichkeit getan werden kann (vgl. Grzybek 2000b).

Tab. 8 gibt die Ergebnisse der Vergleiche zwischen den Studien mit identischer Sprache wieder.

Tab. 8: Mittelwertvergleiche zwischen Stichproben mit Sprichwörtern identischer Sprache

	Stichprobe 1	Stichprobe 2	t	FG	p
Estnisch	Normann Normann 17./18. Jhd.	17./18. Jhd. geläufige geläufige	0.01	874	>0.1
			6.48	349	< 0.001***
			5.56	580	< 0.001***
Russisch	Ivanov	Tóthné Litovkina	2.45	205	< 0.05*
Deutsch	Senaltan Senaltan Senaltan Senaltan Senaltan	Ivanov Simrock (ges.) Simrock (> 50) Simrock (> 75) Simrock (100)	7.72	336	< 0.001***
			1.66	211	> 0.05
			7.99	317	< 0.001***
			8.44	404	< 0.001***
			5.71	102	< 0.001***
	Ivanov Ivanov Ivanov Ivanov	Simrock (ges.) Simrock (> 50) Simrock (> 75) Simrock (100)	12.19	976	< 0.001***
			0.23	1534	> 0.05
			1.76	755	> 0.05
			1.48	73	> 0.05

Wie die in Tab. 8 dargestellten Ergebnisse klar zeigen, unterscheiden sich die Ergebnisse für die verschiedenen Stichproben mit Sprichwörtern ein und derselben Sprache nicht unerheblich voneinander. Dies gilt auf jeden Fall für die Stichproben von Krikmann, im Hinblick auf die sich die vom Autor gezogenen Schlußfolgerungen bestätigen: Während der Unterschied zwischen den Sprichwörtern des 17./18. Jhds. und denen aus der Sammlung von Normann kein Signifikanzniveau erreicht, unterscheiden sich sowohl die Normann-Sprichwörter als auch die des 17./18. Jhds. hoch signifikant von den geläufigen.

Signifikante Unterschiede gibt es auch zwischen den beiden Untersuchungen zu russischen Sprichwörtern von Ivanov und Tóthné Litovkina ( $p < 0.05$ ) sowie zu deutschen Sprichwörtern von Senaltan und Ivanov ( $p < 0.001$ ). Nach einer Begründung für diese Unterschiede ist bislang – außer von Krikmann (1967) und Grzybek (1995) – nicht gefragt worden. Während Krikmann historisch argumentiert und in

einem um das 19. Jhd. noch nicht abgeschlossenen Kodifizierungsprozeß verankert sieht, hat Grzybek (1995) die Vermutung vorgebracht, daß die Unterschiede mit der unterschiedlichen Bekanntheit des zugrunde gelegten Sprichwortmaterials zusammenhängen. Schaut man sich die Ergebnisse der Simrock-Analysen von Grzybek differenziert nach den unterschiedenen Bekanntheitsklassen an, so entsteht in der Tat ein entsprechender Eindruck:

a) Die durchschnittliche Länge der Sprichwörter von Senaltan unterscheidet sich nicht signifikant ( $p > 0.05$ ) von den Simrock'schen Sprichwörtern, wenn man hier das gesamte Korpus als Vergleichsmaterial heranzieht; mit jeder der unterschiedenen Bekanntheitsklassen als Vergleichsbasis hingegen erweisen sich die Längenunterschiede als hoch signifikant ( $p < 0.001$ ).

b) Die durchschnittliche Länge der Sprichwörter von Ivanov hingegen scheint genau dem gegenteiligen Trend zu folgen: Hier unterscheidet sich die durchschnittliche Länge nur, wenn man das gesamte Simrock'sche Korpus zum Vergleich heranzieht, während in den unterschiedenen Bekanntheitsklassen kein Signifikanzniveau erreicht wird ( $p > 0.05$ ).

Es liegt insofern die von Grzybek (1995, 213f.) vorgeschlagene Schlußfolgerung nahe, daß das von Senaltan verwendete Material eher einer undifferenzierten Sprichwortmenge entspricht, während die Einheiten der Untersuchung von Ivanov eher einem Korpus bekannterer Sprichwörter entsprechen. Diese Schlußfolgerung ist allerdings nicht unproblematisch, und zwar aus mindestens zwei Gründen: Erstens – und das ist eigentlich der weniger gewichtige Grund – wird sich diese Hypothese aufgrund fehlender Angaben zur Bekanntheit der Sprichwörter in den Untersuchungen von Ivanov und Senaltan kaum überprüfen lassen. Und zweitens – und das ist ein wesentlich wichtigerer, weil grundsätzlich methodologischer Einwand – ist die Art und Weise, wie der Zusammenhang zwischen Bekanntheit und Sprichwortlänge bei Grzybek (1995) bestimmt wurde,

fragwürdig. Da diese Fragwürdigkeit grundsätzlichen methodologischen Charakters ist, soll sie hier ausführlicher behandelt werden, da sich aus der Problematik weiterführende Perspektiven ableiten lassen.

Wenn man, um eine willkürlich motivierte Datenzusammenfassung herauszugreifen, die Sprichwortmenge der 133 Sprichwörter aus Simrock, die 19-20 der 20 befragten Personen bekannt waren, berechnet, so ergibt sich eine mittlere Länge von  $x = 6.08$  mit einer Standardabweichung von  $s = 2.18$ . Der Unterschied zu den 200 Sprichwörtern von Senaltan erweist sich, wie zu erwarten ist, als hoch signifikant ( $t_{FG-303} = 8.10; p < 0.001$ ). Entgegen der Erwartung stellt sich allerdings auch der Unterschied zu den Sprichwörtern von Ivanov als signifikant dar ( $t_{FG-207} = 2.59; p < 0.05$ ), was im Ergebnis zu dem widersprüchlich scheinenden Gesamtbild führt, daß es keinen Mittelwertunterschied gibt, wenn man die Ivanov'schen Sprichwörter mit den Simrock'schen über 75% Bekanntheit oder mit 100% Bekanntheit vergleicht, wohl aber mit den Simrock'schen über 90% Bekanntheit (s.o.).

De facto stellt sich damit die Vermutung von Grzybek (1995, 214) als richtig heraus, daß es bei der Untersuchung des Zusammenhangs von Sprichwortlänge und Bekanntheit nicht sinnvoll ist, Mittelwertvergleiche anzustellen, die auf mehr oder weniger willkürlich zusammengefaßten Daten beruhen. Angebrachter wäre vielmehr die Berechnung eines Zusammenhangsmaßes oder eine geeignete Kurvenanpassung. Mit diesem Vorgehen würde auch ein anderes methodologisches Problem beseitigt, nämlich die von Grzybek (1995) ohne Bonferoni-Korrektur durchgeführten multiplen Mittelwertvergleiche mit Hilfe von  $t$ -Tests innerhalb ein und derselben Stichprobe (s.o.); erschwerend kommt hinzu, daß die unterschiedenen Untergruppen der Stichprobe sich bei Grzybek (1995) überlappen, insofern die Sprichwörter mit  $> 50\%$  Bekanntheit auch in der Untergruppe mit 100% Bekanntheit enthalten sind, bzw. beide Bekanntheitsklassen sind auch Bestandteil der Gesamtstichprobe.

Die Lösung dieser Fragen kann hier allerdings ebenfalls nicht im Detail weiter verfolgt werden und muß ebenfalls andernorts in eigenem Kontext wieder aufgegriffen werden (Grzybek 2000b).

#### 4. Programm

Aus der obigen Diskussion zum Status von Satzlängenforschungen im Rahmen der Parömiologie leitet sich somit die Schlußfolgerung ab, daß wir noch weit davon entfernt sind, einigermaßen zuverlässige Angaben über sprichwörtliche Satzlängen und die hierbei ins Spiel kommenden Faktoren und Parameter zu treffen. Andererseits ist die Datenlage (zumindest im Anschluß an die obigen Reparaturmaßnahmen) nicht so schlecht, als daß sich auf dieser Basis nicht das folgende Minimalprogramm abzeichnete, nach dessen Absolvierung uns erstmals valide Aussagen zur Verfügung stehen dürften:

- Zunächst gilt es den Faktor der Bekanntheit bzw. dessen Einfluß auf die (durchschnittliche) Sprichwortlänge zu klären. Im Gegensatz zu anderen möglichen Einflußfaktoren (deren Relevanz niemals empirisch-systematisch untersucht wurde) wie z.B. Bildlichkeit, Rhythmus, Metrum u.a. – die als "interne Einflußfaktoren" angesehen werden können –, handelt es sich bei Bekanntheit in gewissem Sinne um einen externen Einflußfaktor, dessen Wirkung es vorrangig zu klären gilt. Durchzuführen sind u.a. Analysen der Schiefe und Wölbung der entsprechenden Verteilungen, in erster Linie allerdings *Regressionsanalysen*. Sollte es sich um eine lineare Abhängigkeit zwischen Bekanntheit und Satzlänge handeln, wäre der Pearson'sche Korrelationskoeffizient zu bestimmen und zur Kontrolle ein Test auf (Abweichung von) Linearität durchzuführen. Im Falle einer nicht-linearen Abhängigkeit wäre ein für nicht-lineare Beziehungen adäquates Zusammenhangsmaß zu bestimmen und eine entsprechende Kurvenanpassung vorzunehmen. Zu untersuchen ist weiterhin die Richtung der Abhängigkeit

(wirkt Bekanntheit auf die Länge oder führt Kürze zu höherer Bekanntheit?), die Auswirkung der zugrundegelegten Meßeinheiten (ist das Wort eine geeignete Größe, wie stellen sich die Ergebnisse dar, wenn man Länge in Silben oder Teilsätzen berechnet), u.a.m. – zu diesem Fragenkomplex s. Grzybek (2000c).

- Durchzuführen sind weiterhin umfassende *Verteilungsanalysen*. Zu stellen sind hier zunächst Fragen wie z.B., ob zwei miteinander verglichene Häufigkeitsverteilungen derselben Grundgesamtheit entstammen oder nicht. Bei der Untersuchung dieser wird sich zeigen, ob Dichotomisierungen von Daten, die mehr oder weniger arbiträr vorgenommen werden (z.B. in Form von Zusammenfassungen in Klassen von 2-7, 4-8 o.ä. Wörtern) sinnvoll sind oder nicht? Konkret durchzuführen wären zunächst  $\chi^2$ -Tests in Form von 2-2-Felder-Tests (bei Dichotomisierung der Daten) oder k'2-Felder-Tests (bei nicht-dichotomisierten Daten). In der weiteren Folge wäre die Frage zu stellen, ob sich die empirisch beobachteten Verteilungen an theoretische Verteilungsmodelle anpassen lassen, ob diese Anpassungen sprachspezifisch oder übersprachlich gültig sind. Zu leisten sein wird weiterhin ein Vergleich zu Satzlängenverteilungen in Fließtexten, und auch bei dieser Fragestellung wird der Faktor der Bekanntheit zu kontrollieren sein (folgen die bekannteren Sprichwörter denselben Verteilungsmodellen wie die weniger bekannten?) – zu diesem Fragenkomplex s. Grzybek (2000b).

- Schließlich gilt es den Zusammenhang zwischen Wort- und Satzlänge in Sprichwörtern systematisch zu untersuchen; zu fragen gilt es, ob und wie sich diese Abhängigkeit übersprachlich theoretisch beschreiben läßt, ob es bei Sprichwörtern verschiedener Sprachen zu identischen oder unterschiedlichen theoretischen Anpassungen kommt und wie diese sich zu den vergleichbaren Abhängigkeiten in Fließtexten verhalten (vgl. Grzybek 2000d).

Ohne Frage wird es im Hinblick auf die Untersuchung von Satzlängen in Sprichwörtern und den sie beeinflussenden bzw. den hier mitwirkenden Faktoren noch weitere Fragen geben – doch die Beantwortung der aufgeführten Fragen kann als Minimalkatalog verstanden werden, an dessen Beantwortung zukünftige Untersuchungen nicht vorbei kommen werden, wenn jemals der Sprung von einer deskriptiven zu einer erklärenden statistischen Beschreibung der Sprichwörter geschafft werden soll.

#### Literatur:

- Grzybek 1995 - Grzybek P. Zur Frage der Satzlänge von Sprichwörtern (unter besonderer Berücksichtigung deutscher Sprichwörter) // Baur, R.S., Chlosta, Ch. (Hrsg.), *Von der Einwortmetapher zur Satzmetapher. Akten des Westfälischen Arbeitskreises Phraseologie / Parömiologie*. Bochum, 1995. 203-217.
- Grzybek 1998 - Grzybek P. Explorative Untersuchungen zur Wort- und Satzlänge kroatischer Sprichwörter. (Am Beispiel der *Poslovice* von Đuro Daničić [1871]) // ПОЛУТРОПОН. К 70-летию В.Н. Топорова. Москва, 1998. С.447-465.
- Grzybek 2000a - Grzybek P. Einleitung // *Quantitative Sprichwortanalysen*. [In Vorb.]
- Grzybek 2000b - Grzybek P. On the distribution of sentence length in proverbs // In Vorb.
- Grzybek 2000c - Grzybek P. Über den Zusammenhang von Bekanntheit und Satzlänge bei Sprichwörtern – ein allgemeines Gesetz // In Vorb.
- Grzybek 2000d - Grzybek P. Der Zusammenhang von Wort- und Satzlänge bei Sprichwörtern – Regel oder Ausnahme von der Regel? // In Vorb.
- Krikmann 1967 - Krikmann A. Keelestatitikat eesti vanasõnadest // *Emakeele seltsi aastaraamat* (13). 127-154.
- Leino 1968 - Leino P. Mitteilung // *Proverbium* (11); 302.
- Sachs 1992 - Sachs L. *Angewandte Statistik. Anwendung statistischer Methoden*. 7., völlig neu bearbeitete Auflage. Berlin u.a., 1992.
- Senaltan 1968 - Senaltan S. *Studien zur sprachlichen Gestalt der deutschen und türkischen Sprichwörter*. Diss., Marburg, 1968.
- Ivanov 1982 - Иванов В.И. О размерах пословиц (на материале немецкого, русского и чувашского языков) // *Сравнительно-сопоставительное изучение фразеологических выражений в процессе обучения иностранному языку*. Чебоксары, 1982. С. 22-34.
- Tóthné Litovkina 1990 - Tóthné Litovkina A. Hungarian and Russian Proverbs: A comparative Analysis // *Proverbium. Yearbook of International Proverb Scholarship*, vol. 7. The University of Vermont, 1990. 241-254.

<sup>1</sup> Die Ursache dieser Differenz könnte in der unterschiedlichen Berechnung zusammengesetzter Wörter begründet sein.

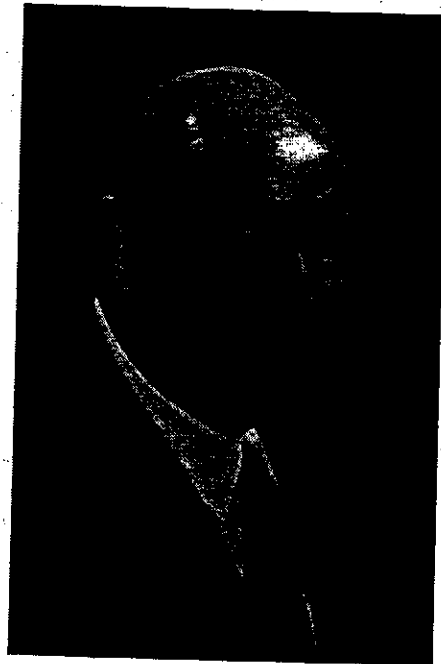
<sup>2</sup> Im Hinblick auf die im vorliegenden Aufsatz angegebenen Zahlen ist zu sagen, daß diese von den in Chlosta/Grzybek/Reos und Grzybek (1995) geringfügig abweichen, ohne daß die gezogenen Schlußfolgerungen davon wesentlich betroffen wären. Die Unterschiede hängen u.a. damit zusammen, daß es in der genannten Sammlung einige Sprichwörter gibt (insgesamt 0.1%), die formal aus mehr als einem Satz bestehen – für die vorliegende Untersuchung wurden die Sprichwörter syntaktisch so bearbeitet, daß ein Sprichwort genau einem Satz entspricht.

<sup>3</sup> Da die dem *t*-Test zugrundeliegende Student-Verteilung bei einem gegen die Normalverteilung konvergiert und bei  $N > \text{ca. } 100$  hinreichend genau die Normalverteilung approximiert, wäre in den meisten Fällen auch ein Normalverteilungstest möglich gewesen. Durchgeführt wurde aber aus Gründen des einheitlichen Vorgehens durchgehend der *t*-Test.

САНКТ-ПЕТЕРБУРГСКИЙ  
ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ

---

Филологический факультет



СЛОВО  
ВО ВРЕМЕНИ  
И  
ПРОСТРАНСТВЕ

*К 60-летию  
профессора В.М.Мокиенко*

Санкт-Петербург  
"ФОЛИО-ПРЕСС"

2000